Network Traffic-Based Malware Detection for Android Devices

Dincy R. Arikkat[1], Hasna Jabeen N. K.[1], Mirko Lai[2], Serena Nicolazzo [2], Antonino Nocera[3], Anafa P.[1], Vinod P.[1], Rafidha Rehiman K.A.[1]

ITADATA, 2025

[1] Department of Computer Applications, Cochin University of Science and Technology

[2] Department of Science and Technological Innovation, University of Eastern Piedmont

[3] Department of Electrical, Computer and Biomedical Engineering, University of Pavia

DIPARTIMENTO DI SCIENZE E INNOVAZIONE TECNOLOGICA

UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE

Overview

1. Motivations

2. Contributions

3. Proposed Approach

4. Experiment Results

5. Conclusion and Future Work

01

1. Motivations

UNIVERSITÀ DEL PIEMONTE ORIENTALE

Android Malware Detection

- The open architecture of the Android system makes it especially vulnerable to malicious apps, posing significant threats to user privacy and security.
- Researchers continually seek innovative approaches to improve mobile malware detection methodologies.

02

2. Contributions

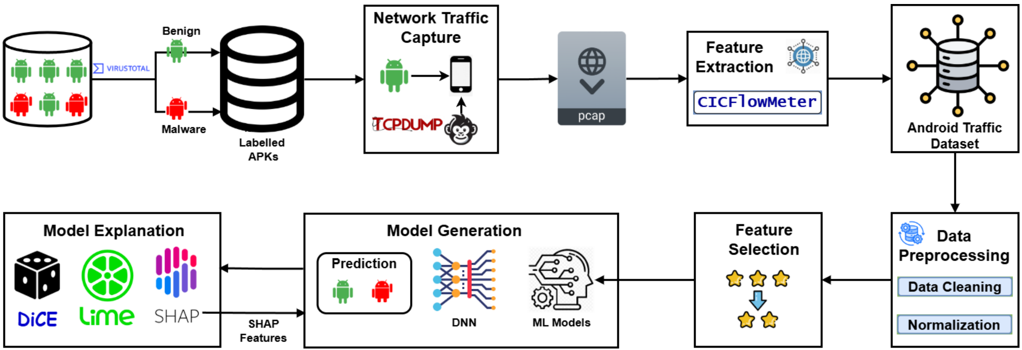
Contributions

- We captured and analyzed mobile network traffic using TCPdump3.
- We evaluated six distinct ML models, and the XGBoost classifier was ultimately selected for its superior performance.
- We explained the predictions using SHapley Additive exPlanations (SHAP) values, Local Interpretable Model-agnostic Explanations (LIME), and Counter Factuals.
- We investigated the potential of SHAP for identifying important features and acting as a feature selection mechanism.

03

3. Proposed Approach

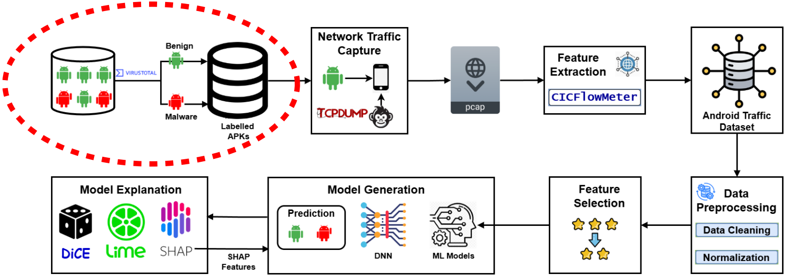
Proposed Approach



UNIVERSITÀ DEL PIEMONTE ORIENTALE

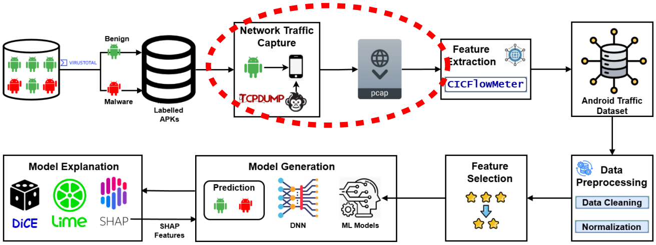
Data Collection

- Androzoo dataset comprises a diverse range of Android applications collected from various sources.
- We selected a total of 8, 819 applications from this dataset.
- To confirm whether these applications were benign or malicious, we evaluated them using VirusTotal.



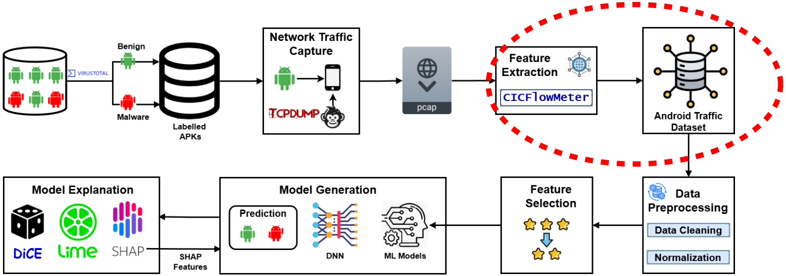
Network Traffic Capture

- We deployed APK files for installation and execution within an Android emulator.
- We employed *Tcpdump* to capture network traffic, a robust command-line packet analyzer tool designed for Linux/Unix systems.
- The Android Monkey tool is utilized to simulate random user interactions.
- Then, the network traffic generated is stored in PCAP files for detailed analysis.
- We generated a Network Traffic APK Dataset comprising 2, 980 malware samples and 5, 866 benign samples.



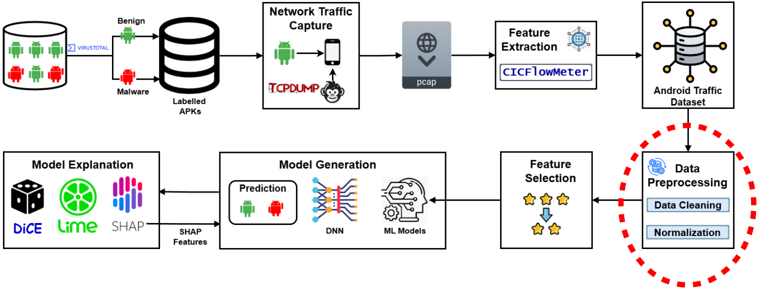
Feature Extraction

- We use *CICFlowMeter*, a portable and scalable open-source tool, to extract a total of 76 relevant features.
- These features include metrics such as flow bytes, total packets in the forward and backward directions, and the mean inter-arrival time of flows, among others.



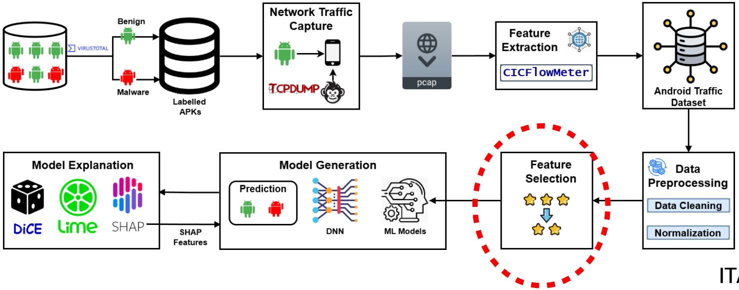
Data Preprocessing

- We identified and addressed errors, missing values, and inconsistent data, which must be corrected or removed before proceeding with the analysis.



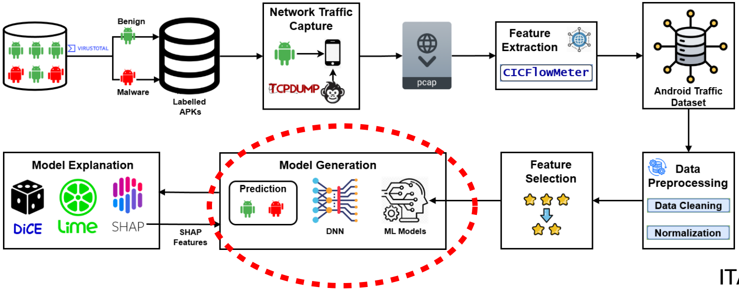
Feature Selection

- We exploited the *SelectKBest* method for feature selection, a univariate feature selection approach.
- This method selects the top *k* features based on their scores, which are computed through the ANOVA F-value, and uses f_classif as the scoring function.
- The *k* best features were calculated for various ML models, including KNN, logistic regression, SVM, decision tree, random forest, and XGBoost.



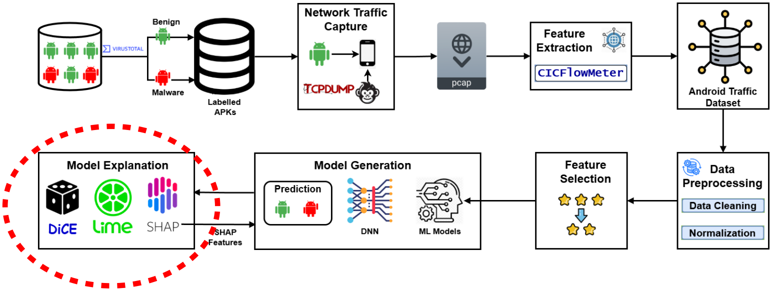
Model Generation

- We performed the *Parameter Tuning* in which we applied *RandomizedSearchCV* to tune these hyperparameters for different models: Random Forest, XGBoost, SVM, Logistic Regression, KNN, and Decision Tree.
- then, in the *Model Training*, we applied several ML and DL techniques to classify and predict malware, including Decision Trees, Random Forest, Support Vector Machines, Logistic Regression, K- Nearest neighbor, XGBoost, and DNN.



Model Explanation

- We exploited SHAP, LIME, and CounterFactuals to provide interpretability of our findings.
- Finally, we conducted an additional round of feature selection using SHAP, leveraging its ability to provide global explanations and identify the most important features contributing to predictions.



UPO UNIVERSITÀ DEL PIEMONTE ORIENTALE

04

4. Experiment Results

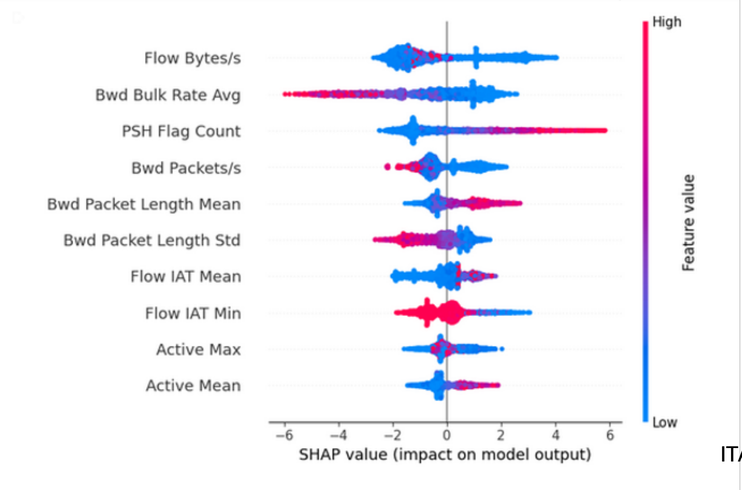
Comparison of ML and DL models

Model	Precision	Recall	F1-Score
Logistic Regression	0.7870	0.7732	0.7772
KNN	0.8211	0.8237	0.8184
Decision Tree	0.8527	0.8515	0.8520
SVM	0.8715	0.8656	0.8673
DNN	0.8944	0.8951	0.8938
Random Forest	0.9087	0.9093	0.9085
XGBoost	**0.9323**	**0.9325**	**0.9323**

Table: Performance Comparison of Android Malware Detection Models using network traffic

Comparison of XAI

Correlation between specific features and the prediction outcome for Android malware detection.



Lime Explanation

Features `Packet Length Variance`, `Bwd Packet Length Std`, `Flow packets` and `Fwd Packets Length Std` contribute more to the Benign class.

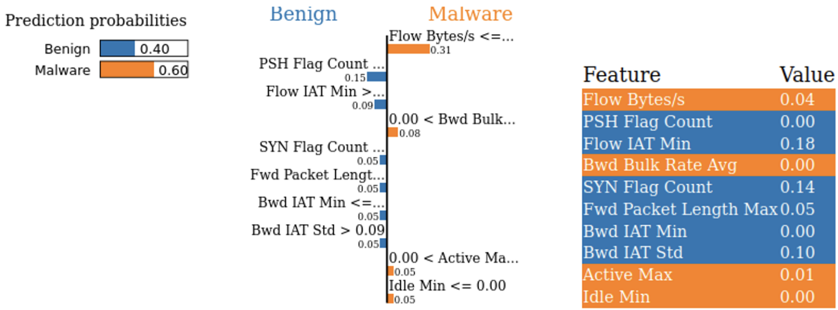


Figure: Lime explanation of Malware class

Frame Title

Features `Flow Bytes`, `Bwd Bulk Rate Avg`, `Active Max`, and `Idle Min` contribute more to the Malware class.

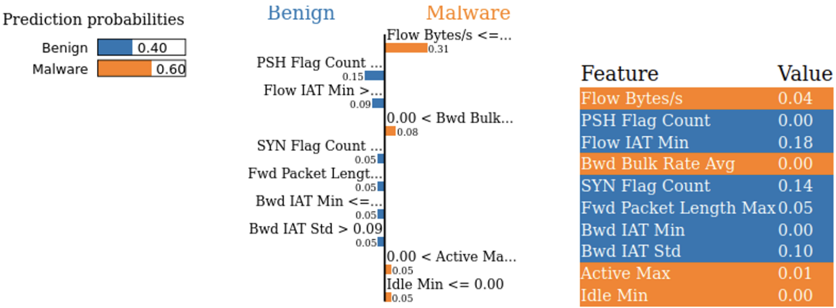


Figure: Lime explanation of Malware class

Feature selection based on the SHAP approach

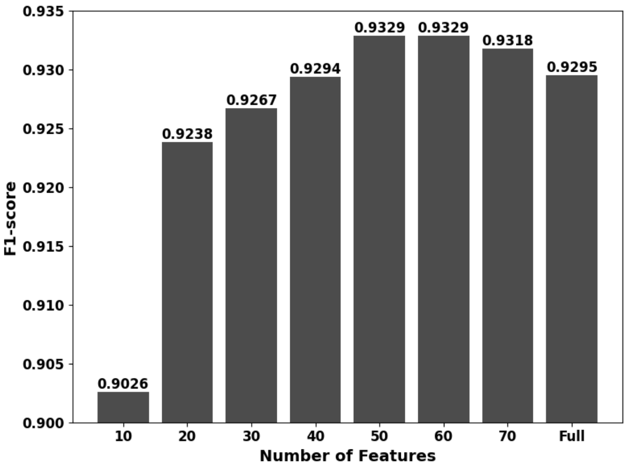


Figure: F1-scores of the models using the SHAP-based features

05

5. Conclusion and Future Work

Conclusion and Future Work

- We collected a dataset of malicious and benign APKs to study the behavior of Android malware.
- By applying several ML and DL techniques, we made significant strides in classifying APKs based on network traffic.
- Our experiments demonstrated that the XGBoost classifier outperformed other algorithms, achieving an impressive weighted F1-score of 93.23% for network traffic analysis.
- We also provide an explainer to help interpret the model.
- Our findings contribute to the advancement of security solutions aimed at reducing the risks associated with Android malware.
- In the future, we plan to integrate additional features such as dynamic analysis and permissions-based data from APKs.

Thanks for the attention!

UNIVERSITÀ DEL PIEMONTE ORIENTALE